

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 30, 2002, 16:15:27 ; Search time 29 Seconds

(without alignments)
461.830 Million cell updates/sec

Title: US-09-664-326-23

Perfect score: 368
Sequence: 1 LKYPCTESGQNLCEGSGN.....PKQSHNDGFERIPREYIQ 65

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	247	67.1	84	5	007557	007557 hirudinaria
2	230	62.5	84	5	007558	007558 hirudinaria
3	138	37.5	25	5	09TWX5	09TWX5 hirudinaria
4	136	37.0	25	5	09TWX4	09TWX4 hirudinaria
5	74.5	20.2	1331	13	090W93	090W93 poecilia re
6	74	20.1	248	5	096962	096962 suberites d
7	74	20.1	2352	5	061240	061240 halocynthia
8	71	19.3	608	5	09VW80	09VW80 drosophila
9	70	19.0	169	2	08RP58	08RP58 anaplasma p
10	69	19.0	323	2	093NT8	093NT8 anaplasma p
11	69	18.8	315	13	09W600	09W600 gallus gall
12	68.5	18.6	2233	5	094711	094711 paramacium
13	68	18.5	184	2	09RB30	09RB30 anaplasma p
14	68	18.5	278	2	09XCP8	09XCP8 anaplasma p
15	68	18.5	364	2	084968	084968 anaplasma p
16	66.5	18.1	516	3	0960Y3	0960Y3 saccharomyc

17	66.5	18.1	645	5	09VWR2	09VWR2 drosophila
18	66.5	18.1	762	13	042373	042373 brachydanio
19	66	17.9	761	4	060477	060477 homo sapien
20	66	17.9	1193	13	090819	090819 gallus gall
21	66	17.9	2150	5	044131	044131 caenorhabdi
22	65.5	17.8	213	5	09CXY8	09CXY8 plasmodium
23	65.5	17.8	248	5	096961	096961 geodia cydo
24	65.5	17.8	455	3	08XOR2	08XOR2 neurospora
25	65.5	17.8	594	5	024970	024970 giardia lam
26	65.5	17.8	915	4	09BXN9	09BXN9 homo sapien
27	65.5	17.8	2601	4	09BX84	09BX84 homo sapien
28	65.5	17.8	2601	4	09BX85	09BX84 homo sapien
29	65.5	17.8	2624	4	09BX85	09BX85 homo sapien
30	65.5	17.8	2624	4	09BL87	09BL87 homo sapien
31	65.5	17.8	2689	4	09BX87	09BX87 homo sapien
32	65.5	17.8	2689	4	09BX85	09BX85 homo sapien
33	65.5	17.8	2721	4	09BX83	09BX83 homo sapien
34	65.5	17.8	2721	4	09BX90	09BX90 homo sapien
35	65.5	17.8	2724	4	09BX88	09BX88 homo sapien
36	65.5	17.8	2724	4	09BL89	09BL89 homo sapien
37	65.5	17.8	2812	4	09BX86	09BX86 homo sapien
38	65.5	17.8	2812	4	09BL88	09BL88 homo sapien
39	65	17.7	558	5	017797	017797 caenorhabdi
40	65	17.7	760	11	0925T8	0925T8 ratu
41	65	17.7	760	11	0920P3	0920P3 mus musculu
42	65	17.7	761	11	09QXL0	09QXL0 mus musculu
43	65	17.7	1043	5	017644	017644 caenorhabdi
44	65	17.7	1101	5	0964D2	0964D2 entamoeba h
45	65	17.7	1661	5	077244	077244 chlorophydra

ALIGNMENTS

RESULT 1
ID 007557 PRELIMINARY; PRT; 84 AA.
AC 007557;
DT 01-NOV-1998 (TREMURel. 08, Created)
DT 01-NOV-1998 (TREMURel. 08, Last sequence update)
DT 01-JUN-2001 (TREMURel. 17, Last annotation update)
DE Hirudin HM2 precursor (Bufrudin).
GN HM2.
OS Hirudinaria manliensis (Buffalo leech).
OC Eukaryota; Metazoa; Annelida; Clitellata; Hirudinea;
OC Ancyrocephallida; Hirudiniformes; Hirudinidae; Hirudinaria.
OX NCBI_TaxID=6419;
[1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 21-84.
RX MEDLINE=93285156; PubMed=7685281;
RA Seacherni E., Nitti G., Valsasina B., Orsini G., Visco C., Ferreira M.,
RA Sawyer R.T., Sarmiento P.;
RT "Novel hirudin variants from the leech Hirudinaria manliensis. Amino
RT acid sequence, cDNA cloning and genomic organization.";
RL Eur. J. Biochem. 214:295-304(1993).
CC -!- FUNCTION: HIRUDIN IS A POTENT THROMBIN-SPECIFIC PROTEASE
CC INHIBITOR. IT FORMS A STABLE NON-COVALENT COMPLEX WITH ALPHA-
CC THROMBIN, THEREBY ABOLISHING ITS ABILITY TO CLEAVE FIBRINOGEN.
CC -!- SIMILARITY: BELONGS TO THE HIRUDIN FAMILY.
CC EMBL: X72786; CA51293.1; -;
DR HSSP; P01050; SHIR.
DR InterPro; IPR000429; Hirudin.
DR Pfam; PF00713; Hirudin.1.
DR PRINTS; PR00777; HIRUDIN.
DR PRODOM; PD004216; Hirudin.1.
KW Serine protease inhibitor; Multigene family; Signal.
FT SIGNAL 1 20
FT CHAIN 21 84 HIRUDIN HM2.
FT DISULFID 26 34 BY SIMILARITY.
FT DISULFID 36 48 BY SIMILARITY.
FT DISULFID 42 57 BY SIMILARITY.
SQ SEQUENCE 84 AA; 9004 MW; CCBASDBSE7IBAF07 CRC64;


```

QY      3 YTDCTESGONL---CLCEGSNNCGGQGNKCILGSGEKNQCVTGEGTPKPQSHNDG---DF 56
      ||| | : : : | :| : : | |||| | : | | : ||

```

CG10089.
Drosophila melanogaster (Fruit fly).
OS

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN NCBI_TaxID=7227;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA Gutter C.R., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.H.C., Blaise R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brothstein P., Brotter P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferlita S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jajala M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon R., Nusskern D.R., Paolel J.M.,
 RA Palazotto M., Plitman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Schebler F., Shen H.,
 RA Shue B.C., Siden-Kimms I., Simpson M., Skupski M.P., Smith T.,
 RA Spletter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003537; AAF49810.1; -
 DR HSSP: Q16828; IMKP.
 DR FlyBase: FBgn0036369; CG10089.
 DR InterPro: IPR000340; DS_phosphatase.
 DR InterPro: IPR000387; TYR_phosphatase.
 DR Pfam: PF00782; DSPC; 1.
 DR SMART: SM00195; DSPC; 1.
 DR PROSITE: PS00056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE: PS00054; TYR_PHOSPHATASE_DUAL; 1.
 SQ SEQUENCE 608 AA; 6963 MW; 71EA1355C7660D1 CRC64;

Query Match 19.3%; Score 71; DB 5; Length 608;
 Best Local Similarity 31.8%; Pred. No. 2.3;
 Matches 14; Conservative 10; Mismatches 18; Indels 2; Gaps 1;

OY 11 ONLCLCEGSSNVCGGNKCLGSDGKNCVWGEGTPKPSHNDGDE-- 54
 DB 199 QNRDICEGN--CSRGEKCPGTADVDGYNIGGEGEEDGEGEG 240

RESULT 9
 ID Q8RP58 PRELIMINARY; PRT; 169 AA.
 AC Q8RP58;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Major surface protein 2 hypervariable region (Fragment).
 GN MS2.
 OS Anaplasma phagocytophilum (Ehrlichia phagocytophila).
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettiales;
 OC Anaplasmataceae; Anaplasma.
 RN NCBI_TaxID=948;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BDS;
 RX MEDLINE=21843110; PubMed=11854205;
 RA Caspersen K., Park J.H., Patil S., Dunlter J.S.;
 RT "Genetic variability and stability of Anaplasma phagocytophila msp2
 (p44)."
 RL Infect. Immun. 70:1230-1234(2002).
 DR EMBL: AF443418; AAL77334.1; -
 FT NON_TER 1
 FT NON_TER 1
 SQ SEQUENCE 169 AA; 17639 MW; AB122A72A8F2CB70 CRC64;

Query Match 19.0%; Score 70; DB 2; Length 169;
 Best Local Similarity 29.6%; Pred. No. 0.77;
 Matches 21; Conservative 10; Mismatches 26; Indels 14; Gaps 3;

OY 6 CTESGONLCLCEGSSNVCGGN-----KCLIGSGEKN-QCVTGSGTPKPSHNDGDE-- 56
 DB 185 GDNSSSTTTTSGTNVSETGVFRDFIRATLKEDGSKNWPSSGTPKPTNNAKAVA 123

OY 57 -----EEIPEE 62
 DB 124 KDVLQELTPEE 134

RESULT 10
 ID Q93NY8 PRELIMINARY; PRT; 323 AA.
 AC Q93NY8;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Major surface protein (Fragment).
 OS Anaplasma phagocytophilum (Ehrlichia phagocytophila).
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettiales;
 OC Anaplasmataceae; Anaplasma.
 RN NCBI_TaxID=948;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WI 1;
 RX MEDLINE=21320744; PubMed=11427556;
 RA Lodes M.J., Mohamath R., Reynolds L.D., McNeill P., Kolbert C.P.,
 RA Brulinsma E.S., Benson D.R., Hofmeister E., Reed S.G., Houghton R.L.,
 RA Persing D.H.;
 RT "Serodiagnosis of human granulocytic ehrlichiosis by using novel
 RT combinations of immunoreactive recombinant proteins."
 RL J. Clin. Microbiol. 39:2466-2476(2001).
 DR EMBL: AF356509; AAK69699.1; -
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1.
 FT NON_TER 1
 FT NON_TER 1
 SQ SEQUENCE 323 AA; 34533 MW; 51E0646ED4F8FDED CRC64;

Query Match 19.0%; Score 70; DB 2; Length 323;
 Best Local Similarity 29.6%; Pred. No. 1.5;
 Matches 21; Conservative 10; Mismatches 26; Indels 14; Gaps 3;

OY 6 CTESGONLCLCEGSSNVCGGN-----KCLIGSGEKN-QCVTGSGTPKPSHNDGDE-- 56
 DB 185 GDNSSSTTTTSGTNVSETGVFRDFIRATLKEDGSKNWPSSGTPKPTNNAKAVA 244

OY 57 -----EEIPEE 62
 DB 245 KDVLQELTPEE 255

```

RESULT 11
ID 09w600 PRELIMINARY; PRT; 315 AA.
AC 09w600;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Flk protein.
GN FLK.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN 11
RP SEQUENCE FROM N.A.
RA Patel K.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RX MEDLINE=96427388; PubMed=8812133;
RA Patel K., Connolly D., Anthor H., Nose K.;
RT "Cloning and early dorsal-axial expression of Flk, a chick
RT follistatin-related gene: Evidence for involvement in
RT dorsalisation/neutral induction. .";
RL Dev. Biol. 178:327-342(1996).
DR EMBL; AJ238977; CAB42968.1; -.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR003645; FOLN.
DR Pfam: PF00050; Kazal; 1.
DR SMART; SM00274; FOLN; 1.
DR SMART; SM00280; KAZAL; 1.
SQ SEQUENCE 315 AA; 35815 MW; C01889E05658A67 CRC64;

Query Match 18.8%; Score 69; DB 13; Length 315;
Best Local Similarity 33.3%; Pred. No. 2;
Matches 21; Conservative 6; Mismatches 26; Indels 10; Gaps 3;

QY 7 TESGONLCLC-----EGSNVCGGNGKNCILG-SPDEKNOCVTGTPKPSHNDGDFEEI 59
DB 45 TERGEPCLCIEOCKPHGRVCGSGNGKTYLNHCEHLHDACTIGS---KIQVDYDGCKEK 101
QY 60 PEE 62
DB 102 KSE 104

RESULT 12
ID 094711 PRELIMINARY; PRT; 2233 AA.
AC 094711;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 51C surface protein.
OS Parametium tetraurelia.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Pericollida;
OC Parametium.
OX NCBI_TaxID=5888;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=92106337; PubMed=1762150;
RA Nielsen E., You Y., Forney J.;
RT "Cysteine residue periodicity is a conserved structural feature of
RT variable surface proteins from Parametium tetraurelia. .";
RL J. Mol. Biol. 222:835-841(1991).
DR EMBL; M65164; AAA61740.1; -.
DR InterPro: IPR002895; Parametium_SA.
DR Pfam; PF01508; Parametium_SA; 25.
SQ SEQUENCE 2233 AA; 237078 MW; C064FE0AF7BB873B CRC64;

Query Match 18.6%; Score 68.5; DB 5; Length 2233;

```

```

Best Local Similarity 39.5%; Pred. No. 19;
Matches 17; Conservative 4; Mismatches 13; Indels 9; Gaps 3;

QY 6 CTESGONLCLCEGSGNVCGGNGN---KCIIGSGENK-QCVTEGTPKPSHNDGDFE-- 56
DB 1830 CTIDGGO---CIGITGCGKTNVNGCVTGIDGE---CITTVAT 1866

RESULT 13
ID 09RB30 PRELIMINARY; PRT; 184 AA.
AC 09RB30;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 44 kDa major outer membrane protein (Fragment).
GN P44-15.
OS Anaplasma phagocytophilum (Ehrlichia phagocytophila).
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Anaplasma.
OX NCBI_TaxID=948;
RN 11
RP SEQUENCE FROM N.A.
RX STRAIN=HZ;
RL MEDLINE=99292751; PubMed=10364227;
RA Zhi N., Ohashi N., Rikihisa Y.;
RT "Multiple p44 genes encoding major outer membrane proteins are
RT expressed in the human granulocytic ehrlichiosis agent. .";
RL J. Biol. Chem. 274:17828-17836(1999).
DR EMBL; AF135261; AAD41485.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 184 AA; 19245 MW; 2BC9AD4775FD55A1 CRC64;

Query Match 18.5%; Score 68; DB 2; Length 184;
Best Local Similarity 29.6%; Pred. No. 1.5;
Matches 21; Conservative 10; Mismatches 26; Indels 14; Gaps 3;

QY 6 CTESGONLCLCEGSGNVCGGNGN---KCIIGSGENK-QCVTEGTPKPSHNDGDFE-- 56
DB 77 CGDNGSSSTITNSGANVSETVGFDFRATRLKEDGSKNWPSTSGTGPVYNDNAKAVA 136
QY 57 -----EIPPEE 62
DB 137 KDLYQELTPPE 147

RESULT 14
ID 09XCP8 PRELIMINARY; PRT; 278 AA.
AC 09XCP8;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 44 kDa major outer membrane protein (Fragment).
GN P44-15.
OS Anaplasma phagocytophilum (Ehrlichia phagocytophila).
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Anaplasma.
OX NCBI_TaxID=948;
RN 11
RP SEQUENCE FROM N.A.
RX STRAIN=HZ;
RL MEDLINE=99292751; PubMed=10364227;
RA Zhi N., Ohashi N., Rikihisa Y.;
RT "Multiple p44 genes encoding major outer membrane proteins are
RT expressed in the human granulocytic ehrlichiosis agent. .";
RL J. Biol. Chem. 274:17828-17836(1999).
DR EMBL; AF135261; AAD41480.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.

```

